PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 9/04, 15/63, 15/86, 15/67, 15/82 — A1

(11) International Publication Number: WO 00/34448

(43) International Publication Date: 15 June 2000 (15.06.00)

(21) International Application Number: PCT/US99/28616

(22) International Filing Date: 3 December 1999 (03.12.99)

(30) Priority Data:
60/110,865 4 December 1998 (04.12.98) US

(71) Applicant *(for all designated States except US)*: E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). LEE, Jian-Ming [CN/US]; 10 Chestnut Place, Nutley, NJ 07110 (US). TAO, Yong [CN/US]; 101-8 Thorn Lane, Newark, DE 19711 (US).

(74) Agent: BEARDELL, Lori, Y.; E.I. Du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PLANT 1-DEOXY-D-XYLULOSE 5-PHOSPHATE REDUCTOISOMERASE

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding an isopentenyl diphosphate biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the isopentenyl diphosphate biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the isopentenyl diphosphate biosynthetic enzyme in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE

PLANT 1-DEOXY-D-XYLULOSE 5-PHOSPHATE REDUCTOISOMERASE

This application claims the benefit of U.S. Provisional Application No. 60/110,865, filed December 4, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase in plants and seeds.

BACKGROUND OF THE INVENTION

Isoprenoids comprise the largest family of natural products, including numerous secondary compounds which play different functional roles in plants as hormones, photosynthetic pigments, electron carriers, and structural components of membranes. The fundamental unit in isoprenoid biosynthesis, isopentenyl diphosphate (IPP), is normally synthesized by the condensation of acetyl CoA through the mevalonate pathway. In many organisms including several bacteria, algae and plant plastids, IPP is synthesized by a mevalonate-independent pathway. The initial step in this pathway is the condensation of pyruvate and glyceraldehyde 3-phosphate to form 1-deoxy-D-xylulose 4-phosphate. In the committed step towards IPP formation 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzes in a single step an intramolecular rearrangement and reduction of 1-deoxy-D-xylulose 4-phosphate to form 2-C-methyl-D-erythritol 4-phosphate.

The *E. coli* 1-deoxy-D-xylulose 5-phosphate reductoisomerase enzyme has only recently been identified. Comparison of the amino acid sequence of the *E. coli* 1-deoxy-D-xylulose 5-phosphate reductoisomerase with those of *Bacillus subtilis*, *Haemophilus influenzae*, *Helicobacter pyroli*, *Mycobacterium tuberculosis* and *Synechocystis sp.* PCC6803 showed that there is little conservation among these sequences (Takahashi et al. (1998) *Proc. Natl. Acad. Sci. USA 95*:9879-9884).

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding isopentenyl diphosphate biosynthetic enzymes. Specifically, this invention concerns an isolated nucleic acid fragment encoding a 1-deoxy-D-xylulose 5-phosphate reductoisomerase. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase.

The present invention also relates to compositions comprising isolated 1-deoxy-D-xylulose 5-phosphate reductoisomerase polynucleotides.

An additional embodiment of the instant invention pertains to isolated polynucleotides comprising the comprising at least one of 30 contiguous nucleotides of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19.

In another embodiment, the present invention relates to an expression cassette comprising an isolated polynucleotide of the present invention operably linked to a promoter.

The present invention relates to a method for positive selection of a transformed cell comprising:

(a) transforming a host cell with the chimeric gene of the present invention or an expression cassette of the present invention; and

(b) growing the transformed host cell under conditions allowing expression of the polynucleotide in an amount sufficient to complement a 1-deoxy-D-xylulose 5-phosphate reductoisomerase null mutant to provide a positive selection means.

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a first polypeptide of at least 200 amino acids that has at least about 93%, more preferably at least about 95%, and more preferably at least about 98% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a polypeptide of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

It is preferred that the isolated polynucleotides of the claimed invention consist of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20. The present invention also relates to an isolated polynucleotide comprising a nucleotide sequences of at least one of 40 (preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 and the complement of such nucleotide sequences.

The present invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to suitable regulatory sequences.

The present invention relates to an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell. The present invention also relates to a virus, preferably a baculovirus, comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

The present invention relates to a process for producing an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting an isolated compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

The present invention relates to a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide of at least 200 amino acids comprising at least about 93%, more preferably at

least about 95%, and more preferably at least about 98% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12,1 4,16, 18, and 20.

The present invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide in a host cell, the method comprising the steps of: (a) constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; (b) introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; (c) measuring the level a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide in the host cell containing the isolated polynucleotide; and (d) comparing the level of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide in the host cell containing the isolated polynucleotide with the level of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide in a host cell that does not contain the isolated polynucleotide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 40 (preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase amino acid sequence.

The present invention also relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a 1-deoxy-D-xylulose 5-phosphate reductoisomerase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in the transformed host cell; (c) optionally purifying the

1-deoxy-D-xylulose 5-phosphate reductoisomerase expressed by the transformed host cell; (d) treating the 1-deoxy-D-xylulose 5-phosphate reductoisomerase with a compound to be tested; and (e) comparing the activity of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase that has been treated with a test compound to the activity of an untreated 1-deoxy-D-xylulose 5-phosphate reductoisomerase, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWING AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawing and Sequence Listing which form a part of this application.

Figure 1 shows a comparison of the amino acid sequences of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase from corn clone p0004.cb1hh74r (SEQ ID NO:16), rice clone rlr6.pk0073.d5 (SEQ ID NO:6), a soybean contig assembled from clones sml1c.pk001.c15, sml1c.pk005.a24, sl1.pk0021.a6, sl2.pk124.p17, sl1.pk0036.a5, sl2.pk0111.c9, sl1.pk152.i19, and sl2.pk0039.d4 (SEQ ID NO:8), a soybean contig assembled from clones ses2w.pk0029.e5, sgc3c.pk001.d16, and sr1.pk0008.d1:fis (SEQ ID NO:18), wheat clone wlm12.pk0003.d11:fis (SEQ ID NO:20), *Arabidopsis thaliana* (NCBI General Identifier No. 4886307; SEQ ID NO:21), and *Mentha x piperita* (NCBI General Identifier No. 4581856; SEQ ID NO:22). Amino acids conserved among all sequences are indicated with an asterisk (*) on the top row; dashes are used by the program to maximize alignment of the sequences.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1

Isopentenyl Diphosphate Biosynthetic Enzymes

Protein	Clone Designation	SEQ ID NO: (Nucleotide)	SEQ ID NO: (Amino Acid)
Corn 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Contig of: p0004.cb1hh74r p0012.cglac07r p0006.cbyvo28r	1	2
Corn 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Contig of: cen3n.pk0157.e12 cr1n.pk0095.g3 cho1c.pk004.f12 csi1.pk0041.f11	3	4
Rice 1-deoxy-D-xylulose 5-phosphate reductoisomerase	rlr6.pk0073.d5	5	6
Soybean 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Contig of: sml1c.pk001.c15 sml1c.pk005.a24 sl1.pk0021.a6 sl2.pk124.p17 sl1.pk0036.a5 sl2.pk0111.c9 sl1.pk152.i19 sl2.pk0039.d4	7	8
Soybean 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Contig of: sr1.pk0008.d1 sr1.pk0007.c11 srm.pk0014.f8	9	10
Wheat 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Contig of: wlm12.pk0003.d11 wr1.pk0084.a4	11	12
Wheat 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Wlm24.pk0014.d7	13	14
Corn 1-deoxy-D-xylulose 5-phosphate reductoisomerase	p0004.cb1hh74r	15	16
Soybean 1-deoxy-D-xylulose 5-phosphate reductoisomerase	Contig of: ses2w.pk0029.e5 sgc3c.pk001.d16 sr1.pk0008.d1:fis	17	18
Wheat 1-deoxy-D-xylulose 5-phosphate reductoisomerase	wlm12.pk0003.d11:fis	19	20

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res. 13*:3021-3030 (1985) and in the *Biochemical J. 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and

format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, a "polynucleotide" is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least one of 60 contiguous nucleotides, preferably at least one of 40 contiguous nucleotides, most preferably one of at least 30 contiguous nucleotides, of the nucleic acid sequence of the SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid

fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of a polypeptide in a plant cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide, such as 1-deoxy-D-xylulose 5-phosphate reductoisomerase in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial, or viral) may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; measuring the level a polypeptide in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide in the host cell containing the isolated polynucleotide with the level of a polypeptide in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those

skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are at least about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 93% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least about 50 amino acids, preferably at least about 100 amino acids, more preferably at least about 150 amino acids, still more preferably at least about 200 amino acids, and most preferably at least about 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS.* *5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST

(Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards

those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences

have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* *3*:225-236).

The "3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptide by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several isopentenyl diphosphate biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the

BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other 1-deoxy-D-xylulose 5-phosphate reductoisomerases, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA 85*:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA 86*:5673-5677; Loh et al. (1989) *Science 243*:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques 1*:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably one of at least 40, most preferably one of at least 30) contiguous nucleotides derived from a nucleotide sequence selected from the

group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide. The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a polypeptide of a gene (such as 1-deoxy-D-xylulose 5-phosphate reducto-isomerases) preferably a substantial portion of a plant polypeptide of a gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a polypeptide.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol. 36*:1-34; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of plastid IPP in those cells. Because this mevalonate-independent pathway appears to be unique to microorganisms and plant plastids inhibitors of 1-deoxy-D-xylulose 5-phosphate reductoisomerases should have no affect on animals making this enzyme an excellent herbicide candidate. Overexpression of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene will produce the active enzyme for high-throughput screening to find inhibitors for this enzyme. These inhibitors may lead to the discovery of novel herbicides.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J. 4*:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptide to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode the instant polypeptide with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.100*:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In

addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds, and is not an inherent part of the invention. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded 1-deoxy-D-xylulose 5-phosphate reductoisomerase. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 6).

Additionally, the instant polypeptides can be used as a targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the polypeptide described herein catalyzes isopentenyl diphosphate synthesis via the mevalonate-independent pathway. Accordingly, inhibition of the activity of the enzyme described herein could lead to inhibition of plant growth. Accordingly, inhibition of the activity of 1-deoxy-D-xylulose 5-phosphate reductoisomerase could lead to inhibition of plant growth. Thus, the instant polypeptides could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter 4*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet. 7*:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Res. 5*:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med. 11*:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics 16*:325-332), allele-specific ligation (Landegren et al. (1988) *Science 241*:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res. 18*:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet. 7*:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res. 17*:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to

design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA 86*:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA 92*:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean, and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean, and Wheat

Library	Tissue	Clone
cen3n	Corn Endosperm 20 Days After Pollination	cen3n.pk0157.e12
cho1c	Corn Embryo 20 Days After Pollination	cho1c.pk004.f12
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0095.g3
csi1	Corn Silk	csi1.pk0041.f11
p0004	Corn Immature Ear	p0004.cb1hh74r
p0006	Corn Young Shoot	p0006.cbyvo28r
p0012	Corn Middle 3/4 of the 3rd Leaf Blade and Mid Rib From Green Leaves Treated with Jasmonic Acid (1 mg/ml in 0.02% Tween 20) for 24 Hours Before Collection	p0012.cglac07r
rlr6	Rice Leaf 15 Days After Germination, 6 Hours After Infection of Strain *Magaporthe grisea* 4360-R-62 (AVR2-YAMO); Resistant	rlr6.pk0073.d5
ses2w	Soybean Embryogenic Suspension Two Weeks After Subculture	ses2w.pk0029.e5
sgc3c	Soybean Cotyledon 14-21 Days After Germination (Starting to Turn Yellow)	sgc3c.pk001.d16
sl1	Soybean Two-Week-Old Developing Seedlings	sl1.pk0021.a6
sl1	Soybean Two-Week-Old Developing Seedlings	sl1.pk0036.a5
sl1	Soybean Two-Week-Old Developing Seedlings	sl1.pk152.i19
sl2	Soybean Two-Week-Old Developing Seedlings Treated With 2.5 ppm chlorimuron	sl2.pk0039.d4
sl2	Soybean Two-Week-Old Developing Seedlings Treated With 2.5 ppm chlorimuron	sl2.pk0111.c9
sl2	Soybean Two-Week-Old Developing Seedlings Treated With 2.5 ppm chlorimuron	sl2.pk124.p17
sml1c	Soybean Mature Leaf	sml1c.pk001.c15
sml1c	Soybean Mature Leaf	sml1c.pk005.a24
sr1	Soybean Root	sr1.pk0008.d1
srm	Soybean Root Meristem	srm.pk0014.f8
wlm12	Wheat Seedlings 12 Hours After Inoculation With *Erysiphe graminis f. sp tritici*	wlm12.pk0003.d11
wlm24	Wheat Seedlings 24 Hours After Inoculation With *Erysiphe graminis f. sp tritici*	wlm24.pk0014.d7
wr1	Wheat Root From 7 Day Old Seedling	wr1.pk0084.a4

*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA

libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science 252*:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding isopentenyl diphosphate biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet. 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to 1-deoxy-D-xylulose 5-phosphate

reductoisomerase from *Synechocystis PCC6803* and *Escherichia coli* (NCBI General Identifier Nos. 1001556 and 3434984, respectively). Shown in Table 3 are the BLAST results for individual ESTs ("EST"), contigs assembled from two or more ESTs ("Contig"), or sequences encoding the entire protein derived from the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), a contig, or an FIS and PCR ("CGS"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase

Clone	Status	BLAST pLog Score 1001556	BLAST pLog Score 3434984
Contig of: p0004.cb1hh74r p0012.cglac07r p0006.cbyvo28r	Contig	14.40	10.70
Contig of: cen3n.pk0157.e12 cr1n.pk0095.g3 cho1c.pk004.f12 csi1.pk0041.f11	Contig	111.0	59.52
rlr6.pk0073.d5	CGS	164.0	94.0
Contig of: sml1c.pk001.c15 sml1c.pk005.a24 sl1.pk0021.a6 sl2.pk124.p17 sl1.pk0036.a5 sl2.pk0111.c9 sl1.pk152.i19 sl2.pk0039.d4	CGS	154.0	85.50
Contig of: sr1.pk0008.d1 sr1.pk0007.c11 srm.pk0014.f8	Contig	64.40	32.40
Contig of: wlm12.pk0003.d11 wr1.pk0084.a4	Contig	12.70	9.30
wlm24.pk0014.d7	EST	24.70	10.70

Further sequencing of some of the above clones yielded new information. The BLASTX search using the nucleotide sequences from clones listed in Table 4 revealed similarity of the polypeptides encoded by the cDNAs to 1-deoxy-D-xylulose 5-phosphate reductoisomerase from *Arabidopsis thaliana, Mentha x piperita*, and *Synechocystis sp.* (NCBI General Identifier Nos. 4886307, 4581856, and 2496789, respectively). Shown in Table 4 are the BLAST results for the sequences of the entire cDNA inserts comprising the

indicated cDNA clones ("FIS"), contigs assembled from an FIS and an EST ("Contig "), or sequences encoding the entire protein derived from an FIS, or an FIS and PCR ("CGS"):

TABLE 4

BLAST Results for Sequences Encoding Polypeptides Homologous to 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase

Clone	Status	BLAST pLog Score 4886307	4581856	2496789
p0004.cb1hh74r	CGS	>254.00	>254.00	>254.00
Contig of: ses2w.pk0029.e5 sgc3c.pk001.d16 sr1.pk0008.d1:fis	CGS	>254.00	>254.00	>254.00
wlm12.pk0003.d11:fis	FIS	145.00	145.00	145.00

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:6, 8, 16, 18, and 20 and the *Arabidopsis thaliana* and *Mentha x piperita* sequences (SEQ ID NO:21 and SEQ ID NO:22). The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:6, 8, 16, 18, and 20 and the *Arabidopsis thaliana* and *Mentha x piperita* sequences (SEQ ID NO:21 and SEQ ID NO:22).

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase

SEQ ID NO.	Percent Identity to 4886307	4581856
6	90.9	73.8
8	91.6	73.0
16	88.4	74.1
18	77.6	66.1
20	89.7	72.2

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and

probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode one corn, one rice, one wheat, and two soybean 1-deoxy-D-xylulose 5-phosphate reductoisomerase. These sequences represent the first corn, rice, soybean, and wheat sequences encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase.

EXAMPLE 4

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum

of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* *8*:833-839).

EXAMPLE 5

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* *261*:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) *327*:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and

Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 7

Evaluating Compounds for Their Ability to Inhibit the Activity of Isopentenyl Diphosphate Biosynthetic Enzymes

The polypeptides described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 6, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant polypeptides may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("$(His)_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant polypeptides, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant polypeptides are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant polypeptides may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a $(His)_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant polypeptides disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for 1-deoxy-D-xylulose 5-phosphate reductoisomerase are presented by Kuzuyama et al. (1998) *Tetrahedron lett. 39*:4509-4512.

Various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

The disclosure of each reference set forth above is incorporated herein by reference in its entirety.

CLAIMS

What is claimed is:

1. An isolated polynucleotide comprising a first nucleotide sequence encoding a polypeptide of at least 200 amino acids that has at least 93% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20, or a second nucleotide sequence that is complementary to the first nucleotide sequence.

2. The isolated nucleic acid fragment of Claim 1 wherein the first nucleotide sequence consists of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19, that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20.

3. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are DNA.

4. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are RNA.

5. A chimeric gene comprising the isolated polynucleotide of Claim 1 operably linked to suitable regulatory sequences.

6. An isolated host cell comprising the chimeric gene of Claim 5.

7. An isolated host cell comprising an isolated polynucleotide of Claim 1 or Claim 3.

8. The isolated host cell of Claim 7 wherein the isolated host selected from the group consisting of yeast, bacteria, plant, and virus.

10. A virus comprising the isolated polynucleotide of Claim 1.

11. A composition consisting of a polypeptide of at least 200 amino acids that has at least 93% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20.

12. A method of selecting an isolated polynucleotide that affects the level of expression of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide in a host cell, the method comprising the steps of:

(a) constructing an isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such nucleotide sequences;

(b) introducing the isolated polynucleotide into a host cell;

(c) measuring the level of a polypeptide in the host cell containing the polynucleotide; and

(d) comparing the level of polypeptide in the host cell containing the isolated polynucleotide with the level of polypeptide in a host cell that does not contain the isolated polynucleotide.

13. The method of Claim 12 wherein the isolated polynucleotide consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, and 20.

14. A method of selecting an isolated polynucleotide that affects the level of expression of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide in a host cell, the method comprising the steps of:

(a) constructing an isolated polynucleotide of Claim 1;

(b) introducing the isolated polynucleotide into a host cell;

(c) measuring the level of polypeptide in the host cell containing the polynucleotide; and

(d) comparing the level of polypeptide in the host cell containing the isolated polynucleotide with the level of polypeptide in a host cell that does not contain the polynucleotide.

15. A method of obtaining a nucleic acid fragment encoding a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide comprising the steps of:

(a) synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 40 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such nucleotide sequences; and

(b) amplifying a nucleic acid sequence using the oligonucleotide primer.

16. A method of obtaining a nucleic acid fragment encoding the amino acid sequence encoding a 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide comprising the steps of:

(a) probing a cDNA or genomic library with an isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such nucleotide sequences;

(b) identifying a DNA clone that hybridizes with the isolated polynucleotide;

(c) isolating the identified DNA clone; and

(d) sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

17. A method for evaluating at least one compound for its ability to inhibit the activity of an isopentenyl diphosphate biosynthetic enzyme, the method comprising the steps of:

(a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an isopentenyl diphosphate biosynthetic enzyme, operably linked to suitable regulatory sequences;

(b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the isopentenyl diphosphate biosynthetic enzyme encoded by the operably linked nucleic acid fragment in the transformed host cell;

(c) optionally purifying the isopentenyl diphosphate biosynthetic enzyme expressed by the transformed host cell;

(d) treating the isopentenyl diphosphate biosynthetic enzyme with a compound to be tested; and

(e) comparing the activity of the isopentenyl diphosphate biosynthetic enzyme that has been treated with a test compound to the activity of an untreated isopentenyl diphosphate biosynthetic enzyme,

thereby selecting compounds with potential for inhibitory activity.

18. A composition comprising the isolated polynucleotide of Claim 1.

19. A composition comprising the isolated polynucleotide of Claim 10.

20. An isolated polynucleotide of Claim 1 comprising the nucleotide sequence comprising at least one of 30 contiguous nucleotides of a nucleic sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and the complement of such sequences.

21. An expression cassette comprising an isolated polynucleotide of Claim 1 operably linked to a promoter.

22. A method for positive selection of a transformed cell comprising:

(a) transforming a host cell with the chimeric gene of Claim 5 or an expression cassette of Claim 21; and

(b) growing the transformed host cell under conditions suitable for the expression of the polynucleotide in an amount sufficient to complement a 1-deoxy-D-xylulose 5-phosphate reductoisomerase mutant to provide a positive selection means.

23. The method of Claim 22 wherein the plant cell is a monocot (corn, wheat, or rice).

24. The method of Claim 22 wherein the plant cell is a dicot.

FIGURE 1

```
SEQ ID NO:21  AP----------------------------------------------------------
SEQ ID NO:22  MALNLMA-PTEIKTLSFLDSSKSN-YNLNPLKFQGGFAFKRKDSRCTAAKRVHCSAQSQS
SEQ ID NO:16  MAALKASFRGELSAASFLDSSRG-PL----VQHKVDFTFQRKGKRAISLRRTCCSMQQAP
SEQ ID NO:06  MALKVVSFPGDLAAVSFLDSNRGGAF----NQLKVDLPFQTRDRRAVSLRRTCCSMQQAP
SEQ ID NO:18  MMALNISSPAEVKSIFFADSFKSNCLTAK---FSGGFAFKRKERRAASGGRVYCSVQATP
SEQ ID NO:08  MALNLPS-PAQVKPLFFSSN--------NSTKLPGSFSLKRKDSDTTVERRVYCSAAAQS
SEQ ID NO:20  TS----SFSLLELSSGTTRSRRGAAFRPRQHQRKVDLTFQRRDKRAAYLR-TCCSMQQGP
              1                                                         60

                                * * ****  *********** **** *  *  * ******
SEQ ID NO:21  --------------R-QSWDGPKPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNV
SEQ ID NO:22  PPPAWPGRAFPEPGR-MTWEGPKPISVIGSTGSIGTQTLDIVAENPDKFRIVALAAGSNV
SEQ ID NO:16  PP-AWPGRAVAEPGR-RSWDGPKPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNV
SEQ ID NO:06  PP-AWPGRAVVEPGR-RSWDGPKPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNV
SEQ ID NO:18  PPPAWPGRAVPEQGR-KTWDGPKPISIVGSTGSIGTQTLDIVAENPDKFKVVALAAGSNV
SEQ ID NO:08  PPPAWPGTAIPEPSDFKTWDGQKPISVLGSTGSIGTQTLSIVAEFPERFKVVSLAAGSNI
SEQ ID NO:20  PP-AWPGRAVAEPER-RSWEGPKPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNV
             61                                                         120

              ******   ***  *    ***  ** ***       ******** ** **** *  ***
SEQ ID NO:21  TLLADQVRRFKPALVAVRNESLINELKEALADLDYKLEIIPGEQGVIEVARHPEAVTVVT
SEQ ID NO:22  TLLADQ-KAFKPKLVSVKDESLISELKEALAGFEDMPEIIPGEQGMIEVARHPDAVTVVT
SEQ ID NO:16  TLLADQVKTFKPKLVAVRNESLVDELKEALADCEEKPEIIPGEQGVIEVARHPDAVTVVT
SEQ ID NO:06  TLLADQVKTFKPKLVAVRNESLVDELKEALADCDWKPEIIPGEQGVIEVARHPDAVTVVT
SEQ ID NO:18  TLLADQVKRFKPQLVAVRNESLIAELEEALHDVEEKPEIIPGEQGIIEVARHPDAVSVVT
SEQ ID NO:08  TLLADQIKTFKPEVVGLRNESLIDELKEALADVDHKPEIIPGEQGVIEAARHPDATTVVT
SEQ ID NO:20  TLLADQVKTFKPKLVAVRNESLLNELKEALAGCEEMPEIIPGEQGVIEVARHPDAVTVVT
            121                                                         180
```

FIGURE 1

```
               **************************** *** ******* **  ***************
SEQ ID NO:21   GIVGCAGLKPTVAAIEAGKDIALANKETLIAGGPFVLPLANKHNVKILPADSEHSAIFQC
SEQ ID NO:22   GIVGCAGLKPTVAAIEAGKDIALANKETLIAGGPFVLPLAKKHNVKILPADSEHSAIFQC
SEQ ID NO:16   GIVGCAGLKPTVAAIEAGKDIALANKETLIAGGPFVLPLAHKHKVKILPADSEHSAIFQC
SEQ ID NO:06   GIVGCAGLKPTVAAIEAGKDIALANKETLIAGGPFVLPLAQKHKVKILPADSEHSAIFQC
SEQ ID NO:18   GIVGCAGLKPTVAAIEAGKDIALANKETLIAGGPFVLPLAQKHNVKILPADSEHSAIFQC
SEQ ID NO:08   GIVGCAGLKPTVAAIEAGKDIALANKETMIAGAPFVLPLAHKHNIKILPADSEHSAIFQS
SEQ ID NO:20   GIVGCAGLKPTVAAIEAGKDIALANKETLIAGGPFVLPLAHKHNVKILPADSEHSAIFQC
              181                                                         240

               ****  * **   ** ******  *    *  * ******* *  * * *     ** *
SEQ ID NO:21   IQGLPEGALRKIILTASGGAFRDWPVEKLKEVKVADALKHPNWNMGKKITVDSATLF-NK
SEQ ID NO:22   IQGLPEGALRRIILTASGGAFRDLPVEKLKEVKVADALKHSNWNMGKKNTVRLLQLFFNK
SEQ ID NO:16   IQGLSEGALRRIILTASGGAFRDWPVDRLKDVKVADALKHPNWNMGRKITVDSATLF-NK
SEQ ID NO:06   IQGLPEGALRRIILTASGGAFRDWPVDKLKEVKVADALKHPNWNMGKKITVDSATLF-NK
SEQ ID NO:18   IQGLPEGALRRVILTASGGAFRDWPVDKLKDVKVADALKHPNWNMGKKITVDSATLF-NK
SEQ ID NO:08   IQGLPKGALRKILLTGSGGAFREWPAEKMKDIKLADALKHPIWSLGRKITIDSATLF-NK
SEQ ID NO:20   IQGLSEGSLRRVILTASGGAFRDWPVEKLKDVKVADALKHPNWSMGKKITVDSATLF-NK
              241                                                         300

               ***** ******* ** ******  *****  ** **** **** **** * *** ***
SEQ ID NO:21   GLEVIEAHYLFGAEYDDIEIVIHPQSIIHSMIETQDSSVLAQLGWPDMRLPILYTMSWPD
SEQ ID NO:22   GLEVIKAHYLFGAEYDDIEIVIHSPSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPE
SEQ ID NO:16   GLEVIEAHYLFGAEYDDIEIVIHPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPD
SEQ ID NO:06   GLEVIEAHYLFGAEYDDIEIVIHPQSIIHSMIETQDSSVLAQLGWPDMRIPTLYTMSWPD
SEQ ID NO:18   GLEVIEAHYLFGADYDHIEIVIHPQSIIHSMIETQDSSVLAQLGWPDMRLPILYTLSWPD
SEQ ID NO:08   GLEVIEAHYLFGASYDDIEIVIHPQSIIHSLVETXDSSVNAQLGIPDMRLPLLYTLSWPE
SEQ ID NO:20   GLEVIEAHYLFGAEYDDIEIVIHPQSIIHSMIETQDSSVLAQLGWPDMRLPILYTLSWPD
              301                                                         360
```

FIGURE 1

```
              *  *** ******* *   * *  **    *   *   *     ***************
SEQ ID NO:21  RVPCSEVTWPRLDLCKLGSLTFKKPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVEM
SEQ ID NO:22  RVYCSEITWPRLDLCKVD-LPFKKPDNREIPAMDLAYAAWKSRSTMTGVLSAANEKAVEM
SEQ ID NO:16  RIYCSEVTWPRLDLCKLGSLTFRAPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVEL
SEQ ID NO:06  RIYCSEVTWPRLDLCKLGSLTFKAPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVEL
SEQ ID NO:18  RIYCSEVTWPRLDLCKLGSLTFKTPDNVKYPSMNLAFSAGRAGGTMTGVLSAANEKAVEM
SEQ ID NO:08  RIYCSEVTWPRLDLSKYGSLTFYAPDDKKFPSVNLCYAAGRAGGTMTGVLSAANEKAVEM
SEQ ID NO:20  RVYCSEVTWPRLDLCKLGSLTFKAPDNVKYPSVDLAYAAGRAGGTMTGVLSAANEKAVEL
             361                                                         420

              *  *** *   ***** **  *  *    ****** *** ***  **          *
SEQ ID NO:21  FIDEKISYLDIFKVVELTCDKHRNELVTSPSLEEIVHYDLWAREYAANVQLSS-GARPV
SEQ ID NO:22  FIDEKIGYLDIFKVVELTCDKHRSEMAVSPSLEEIVHYDQWARDYAATV-LKSAGLSPA
SEQ ID NO:16  FIDEKISYLDIFKVVELTCNAHRNELVTSPSLEEIVHYDLWARRYAASLQPSS-GLSPV
SEQ ID NO:06  FIDEKIGYLDIFKVVELTCDAHRNELVTRPSLEEIIHYDLWAREYAASLQPST-GLSPV
SEQ ID NO:18  FIDEKISYWNLFKVVELTCEKHQNELVSSPSLEEIIHYDLWARKYAASLQDSSS-FTPI
SEQ ID NO:08  FVEEKISYLDIFKVVELTCQEHQKELVASPSLEEIIHYDQWARQYAASLQKXFKCLNPI
SEQ ID NO:20  FIDEKISYLDIFKVVEMTCDAHRNELVTRPSLEEIIHYDQWARKFAANLQPSSSGRSPV
             421                                                        480

SEQ ID NO:21  ----------------------------HA
SEQ ID NO:22  ----------------------------LV
SEQ ID NO:16  ----------------------------PA
SEQ ID NO:06  ----------------------------PV
SEQ ID NO:18  ----------------------------LA
SEQ ID NO:08  FLTYFRSWGCGGLLATASIFCKCIVGSSIL
SEQ ID NO:20  ----------------------------LA
             481                           510
```

SEQUENCE LISTING

<110> E. I. DU PONT DE NEMOURS AND COMPANY

<120> PLANT 1-DEOXY-D-XYLULOSE 5-PHOSPHATE REDUCTOISOMERASE

<130> BB1297

<140>
<141>

<150> 60/110,865
<151> 1998-DECEMBER-04

<160> 22

<170> Microsoft Office 97

<210> 1
<211> 565
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (5)..(9)

<220>
<221> unsure
<222> (450)

<220>
<221> unsure
<222> (549)

<400> 1
```
aatgnnnnna tcaggctgtt acataggggg gcttgcattg tacacaccca acctggccta  60
gcctacccta ctacactcgt gccgattcgg cacgagcagc gacggtcgcc accaccgctc 120
ccctccctct ccccctcctc gcccagcggc aattaccaca gcctccccag caagccggga 180
tggctgcact caaggcatcg ttccggggtg agctcagcgc cgcttccttc ctcgactcca 240
gcaggggacc tctcgtccag cacaaagtgg attttacgtt tcaaaggaag ggcaaacgag 300
ctatttcact gagaaggaca tgctgttcta tgcaacaggc tccaccacca gcatggcctg 360
ggcgagctgt tgctgagcct ggccggagtc atgggatggc ccaaagccta tctcgattgt 420
tggttcaact ggttccatag gaacacagan attggacatt gttgcggaga atcctgataa 480
gttcagagtt gttgctcttg ctgctggatc caatgtcacg cttctagctg atcaggtcaa 540
aacattcana cctaagttgg ttcgg                                      565
```

<210> 2
<211> 63
<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (25)

<220>
<221> UNSURE
<222> (58)

<400> 2
Ala Trp Pro Glu Ser Trp Asp Gly Pro Lys Pro Ile Ser Ile Val Gly
1 5 10 15

Ser Thr Gly Ser Ile Gly Thr Gln Xaa Leu Asp Ile Val Ala Glu Asn
20 25 30

Pro Asp Lys Phe Arg Val Val Ala Leu Ala Ala Gly Ser Asn Val Thr
35 40 45

Leu Leu Ala Asp Gln Val Lys Thr Phe Xaa Pro Lys Leu Val Arg
50 55 60

<210> 3
<211> 868
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (343)

<220>
<221> unsure
<222> (356)

<220>
<221> unsure
<222> (367)

<220>
<221> unsure
<222> (789)

<220>
<221> unsure
<222> (862)

<400> 3
gatgaattga aagaagcctt ggctgattgc gaagagaagc cagaaattat tcctggggag 60
caaggtgtca tagaagttgc tcgccatcca gatgcagtta cagttgtcac agggatagta 120
ggttgtgcag ggctgaagcc tacagttgct gcaattgaag ctggtaaaga catagcattg 180
gcaaacaaag agacacttat tgcaggtggt ccttttgtgc ttccccttgc acacaaacac 240
aaagtgaaaa ttcttccagc tgattctgag cactctgcaa tatttcagtg tatacaaggc 300
ttgtccgaag gtgcacttcg tcgcattatt ctaactgcat cangtggtgc tttcanggac 360
tggccanttg acaggctgaa agatgtaaaa gttgctgacg ctttaaagca tccaaactgg 420
aatatgggaa ggaagatcac agtagattct gctactttat tcaacaaggg tttagaagtt 480
attgaagcac attatttatt tggtgctgaa tatgatgaca ttgagattgt gattcaccca 540
cagtctatca tacactctat ggttgaaacc caggattcat ctgtcctagc tcagttggga 600
tggccagata tgcggttacc aatcttatac accttatcat ggccagatag gagtcctgag 660
cgctgctaat gagaaggccg tggagttgtt cattgacgag aagattagct acctggacat 720
attcaaggtg gtggagctta catgtaacgc gcatcggaac agctggtaac aaccgtcact 780
ggaggaatng tcattacatc gtggcaagaa tatgcagcat cacaacatct ctggctgagc 840
tgtcctgcat atagtctcac anacttgt 868

<210> 4
<211> 217
<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (115)

<220>
<221> UNSURE
<222> (119)

<220>
<221> UNSURE
<222> (123)

<400> 4
Asp Glu Leu Lys Glu Ala Leu Ala Asp Cys Glu Glu Lys Pro Glu Ile
1 5 10 15

Ile Pro Gly Glu Gln Gly Val Ile Glu Val Ala Arg His Pro Asp Ala
20 25 30

Val Thr Val Val Thr Gly Ile Val Gly Cys Ala Gly Leu Lys Pro Thr
35 40 45

Val Ala Ala Ile Glu Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu
50 55 60

Thr Leu Ile Ala Gly Gly Pro Phe Val Leu Pro Leu Ala His Lys His
65 70 75 80

Lys Val Lys Ile Leu Pro Ala Asp Ser Glu His Ser Ala Ile Phe Gln
85 90 95

Cys Ile Gln Gly Leu Ser Glu Gly Ala Leu Arg Arg Ile Ile Leu Thr
100 105 110

Ala Ser Xaa Gly Ala Phe Xaa Asp Trp Pro Xaa Asp Arg Leu Lys Asp
115 120 125

Val Lys Val Ala Asp Ala Leu Lys His Pro Asn Trp Asn Met Gly Arg
130 135 140

Lys Ile Thr Val Asp Ser Ala Thr Leu Phe Asn Lys Gly Leu Glu Val
145 150 155 160

Ile Glu Ala His Tyr Leu Phe Gly Ala Glu Tyr Asp Asp Ile Glu Ile
165 170 175

Val Ile His Pro Gln Ser Ile Ile His Ser Met Val Glu Thr Gln Asp
180 185 190

Ser Ser Val Leu Ala Gln Leu Gly Trp Pro Asp Met Arg Leu Pro Ile
195 200 205

Leu Tyr Thr Leu Ser Trp Pro Asp Arg
210 215

<210> 5
<211> 1901
<212> DNA
<213> Oryza sativa

<400> 5
acactatgac catgattacg ccaagcgcgc aattaaccct cactaaaggg aacaaaagct 60
ggagctccac cgcggtggcg gccgctctag aactagtgga tcccccgggc tgcaggaatt 120
cggcacgagg tttaaaccag acgtcgagtc gagcattaac tcagtcaggg tggccatggc 180
gctcaaggtc gtctctttcc ccggggactt ggccgcggtc tcattcctcg actccaacag 240
aggaggagct ttcaaccagc tcaaagtgga cctcccgttt caaacgaggg acagaagagc 300
agtttccctg agaaggactt gctgttcaat gcaacaggct ccaccaccag catggcctgg 360
tcgagccgtt gttgaacctg ggaggaggtc atgggatggc cccaagccta tctcaattgt 420
tggctcaacc ggttctattg gcacacagac attggacata gttgcggaga atccagataa 480
attccgggtt gttgctcttg ctgctggctc caatgtgact cttctagctg atcaggtgaa 540
aacattcaaa ccaaagcttg ttgctgtaag aaatgagtca ttagttgatg agctaaagga 600
agccttagct gattgtgatt ggaagccaga aattattcct ggtgagcaag gtgtcataga 660
ggttgctcgc cacccagatg cagttacagt tgttactggg atagtagggt gtgcaggact 720
gaagcctaca gttgctgcaa ttgaagctgg gaaagatata gcattggcga acaaagagac 780
acttattgca ggtggtcctt ttgtgcttcc ccttgcacaa aagcacaaag tgaaaatact 840
tcctgctgat tctgagcact ctgctatatt tcagtgtata caaggcttgc ccgaaggagc 900
acttcgccgc attattttga ctgcatcagg tggtgctttc agggactggc cagttgacaa 960
gttgaaagaa gtaaaagttg ctgatgcttt aaagcacccg aactggaata tggggaagaa 1020
gattactgta gattctgcta cattattcaa caagggttta gaagttattg aagcacatta 1080
tttatttggt gctgaatacg atgacattga aattgtgatc cacccacaat ctatcataca 1140
ctctatgatt gaaacccagg attcatctgt gttggctcaa ctgggatggc cagatatgcg 1200
gataccaacc ttatacacca tgtcttggcc agacagaatc tattgctcag aggtcacctg 1260
gccccgacta gatctttgca agctgggttc actgacattc aaagctcctg acaatgtgaa 1320
atacccgtcg atggatctcg cctatgcagc tggaagagct gggggcacca tgacaggagt 1380
tctgagtgct gctaatgaga aggctgtgga gttgttcatc gatgaaaaga tcgggtacct 1440
ggacatcttc aaggtggtgg agctgacatg cgacgctcat cggaatgagc tagtaacaag 1500
gccatcactg gaggagatca tacattatga tctgtgggcg agggagtatg ctgccagcct 1560
acagccatcc actggcctca gccctgtacc tgtctagtac ttgtagcaat acaaaattac 1620
agtagcattg tacactactg ccgtgccagc tccatgcata gtcagcagct ggccactctc 1680
tagctatatc tagatgcgag agaattttaa ggatgtaaat catgccttca catgaataaa 1740
tcgttcgtcc gtgcgttgtg tattcatgta aattttgacg gatggtcaag taaaaataac 1800
aatggcaaat taatttaggg aaaaaaaaaa aaaaaaaact cgaggggggg cccggtaccc 1860
aattcgccct atagtgagtc gtattacgcg cgctcactgg c 1901

<210> 6
<211> 473
<212> PRT
<213> Oryza sativa

<400> 6
Met Ala Leu Lys Val Val Ser Phe Pro Gly Asp Leu Ala Ala Val Ser
1 5 10 15

Phe Leu Asp Ser Asn Arg Gly Gly Ala Phe Asn Gln Leu Lys Val Asp
20 25 30

Leu Pro Phe Gln Thr Arg Asp Arg Arg Ala Val Ser Leu Arg Arg Thr
35 40 45

Cys Cys Ser Met Gln Gln Ala Pro Pro Pro Ala Trp Pro Gly Arg Ala
50 55 60

Val Val Glu Pro Gly Arg Arg Ser Trp Asp Gly Pro Lys Pro Ile Ser
65 70 75 80

Ile Val Gly Ser Thr Gly Ser Ile Gly Thr Gln Thr Leu Asp Ile Val
85 90 95

Ala Glu Asn Pro Asp Lys Phe Arg Val Val Ala Leu Ala Ala Gly Ser
100 105 110

Asn Val Thr Leu Leu Ala Asp Gln Val Lys Thr Phe Lys Pro Lys Leu
115 120 125

Val Ala Val Arg Asn Glu Ser Leu Val Asp Glu Leu Lys Glu Ala Leu
130 135 140

Ala Asp Cys Asp Trp Lys Pro Glu Ile Ile Pro Gly Glu Gln Gly Val
145 150 155 160

Ile Glu Val Ala Arg His Pro Asp Ala Val Thr Val Val Thr Gly Ile
165 170 175

Val Gly Cys Ala Gly Leu Lys Pro Thr Val Ala Ala Ile Glu Ala Gly
180 185 190

Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile Ala Gly Gly Pro
195 200 205

Phe Val Leu Pro Leu Ala Gln Lys His Lys Val Lys Ile Leu Pro Ala
210 215 220

Asp Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln Gly Leu Pro Glu
225 230 235 240

Gly Ala Leu Arg Arg Ile Ile Leu Thr Ala Ser Gly Gly Ala Phe Arg
245 250 255

Asp Trp Pro Val Asp Lys Leu Lys Glu Val Lys Val Ala Asp Ala Leu
260 265 270

Lys His Pro Asn Trp Asn Met Gly Lys Lys Ile Thr Val Asp Ser Ala
275 280 285

Thr Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala His Tyr Leu Phe
290 295 300

Gly Ala Glu Tyr Asp Asp Ile Glu Ile Val Ile His Pro Gln Ser Ile
305 310 315 320

Ile His Ser Met Ile Glu Thr Gln Asp Ser Ser Val Leu Ala Gln Leu
325 330 335

Gly Trp Pro Asp Met Arg Ile Pro Thr Leu Tyr Thr Met Ser Trp Pro
340 345 350

Asp Arg Ile Tyr Cys Ser Glu Val Thr Trp Pro Arg Leu Asp Leu Cys
355 360 365

Lys Leu Gly Ser Leu Thr Phe Lys Ala Pro Asp Asn Val Lys Tyr Pro
370 375 380

Ser Met Asp Leu Ala Tyr Ala Ala Gly Arg Ala Gly Gly Thr Met Thr
385 390 395 400

Gly Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu Leu Phe Ile Asp
405 410 415

Glu Lys Ile Gly Tyr Leu Asp Ile Phe Lys Val Val Glu Leu Thr Cys
420 425 430

Asp Ala His Arg Asn Glu Leu Val Thr Arg Pro Ser Leu Glu Glu Ile
435 440 445

Ile His Tyr Asp Leu Trp Ala Arg Glu Tyr Ala Ala Ser Leu Gln Pro
450 455 460

Ser Thr Gly Leu Ser Pro Val Pro Val
465 470

<210> 7
<211> 1592
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (993)

<220>
<221> unsure
<222> (1402)

<400> 7
gctggttcaa ctgaggtgat ggctttgaat ttgccttctc ccgcccaagt gaagccctta 60
tttttctctt caaataactc cacaaaactt ccaggtagct tttctttgaa gagaaaagat 120
agtgacacaa cagtagagag acgagtttat tgctctgccg ctgctcaatc accaccacca 180
gcatggccag gaacagctat tcccgagcct tctgatttca agacatggga tgggcaaaaa 240
cctatttctg tcttaggatc tacgggttca attggaactc agacactgag tatagtggct 300
gagttcccag aaagatttaa agttgtgagc cttgctgctg gctctaatat tactcttctt 360
gctgaccaga ttaaaacatt taagcctgaa gttgttggtc ttagaaatga gtctttaatt 420
gatgaactca aagaggcttt ggctgatgtg gatcacaaac ccgaaatcat ccctggagag 480
caaggagtca ttgaggccgc tcgtcaccct gatgccacca ctgtagttac aggcatagtt 540
ggttgtgcag gattaaagcc aacagttgca gcaattgaag cagggaaaga catagcattg 600
gccaacaaag agacaatgat tgcgggagcc cctttgttcc ttcctcttgc tcacaaacat 660
aacataaaaa ttcttcccgc tgattcggaa cattctgcaa tttttcagtc tatccagggg 720
ttgccaaagg gtgcacttag gaaaatcctt ttaactggat caggaggtgc tttcagagaa 780
tggcctgctg aaaagatgaa agatattaag cttgctgatg cattaaagca tcccatatgg 840
agtttgggga gaaaaataac tattgactct gctacccttt tcaataaggg tctagaagta 900
attgaagcac attacttgtt tggagcaagc tatgacgata ttgagattgt tattcatcct 960
caatccatca tacattcctt ggttgaaacg cangattcat ctgttaatgc acagttgggg 1020
atacctgaca tgcgcttacc gctcctttat acattatctt ggccagaaag aatctattgc 1080
tctgaagtaa cttggcctcg tcttgatctt agcaagtatg gttctctaac attttatgca 1140
ccggatgaca agaagtttcc atcggtgaat ctttgctatg ctgcgggacg tgctggaggc 1200
accatgacag gagttcttag tgcagcaaat gagaaagctg tagaaatgtt tgttgaagaa 1260
aagattagtt atctggatat attcaaggtt gtggaactaa cttgtcagga acatcaaaag 1320
gaattagtag catctccgtc actcgaagaa attattcact atgaccaatg ggctcgacaa 1380
tatgctgcta gtctgcaaaa angcttcaag tgtttgaatc ccatatttct gacatatttt 1440
agaagttggg gctgtggtgg attgttggca actgctagca tattttgtaa atgtattgtt 1500
ggttcatcaa tcttgtaaaa tgtaaagggg taagctatat aaagtatatg tactcctaaa 1560
agggtttcaa taaaagttct agcttcaaga aa 1592

<210> 8
<211> 499
<212> PRT
<213> Glycine max

<220>
<221> UNSURE
<222> (325)

<220>
<221> UNSURE
<222> (462)

<400> 8
Met Ala Leu Asn Leu Pro Ser Pro Ala Gln Val Lys Pro Leu Phe Phe
1 5 10 15

Ser Ser Asn Asn Ser Thr Lys Leu Pro Gly Ser Phe Ser Leu Lys Arg
20 25 30

Lys Asp Ser Asp Thr Thr Val Glu Arg Arg Val Tyr Cys Ser Ala Ala
35 40 45

Ala Gln Ser Pro Pro Pro Ala Trp Pro Gly Thr Ala Ile Pro Glu Pro
50 55 60

Ser Asp Phe Lys Thr Trp Asp Gly Gln Lys Pro Ile Ser Val Leu Gly
65 70 75 80

Ser Thr Gly Ser Ile Gly Thr Gln Thr Leu Ser Ile Val Ala Glu Phe
85 90 95

Pro Glu Arg Phe Lys Val Val Ser Leu Ala Ala Gly Ser Asn Ile Thr
100 105 110

Leu Leu Ala Asp Gln Ile Lys Thr Phe Lys Pro Glu Val Val Gly Leu
115 120 125

Arg Asn Glu Ser Leu Ile Asp Glu Leu Lys Glu Ala Leu Ala Asp Val
130 135 140

Asp His Lys Pro Glu Ile Ile Pro Gly Glu Gln Gly Val Ile Glu Ala
145 150 155 160

Ala Arg His Pro Asp Ala Thr Thr Val Val Thr Gly Ile Val Gly Cys
165 170 175

Ala Gly Leu Lys Pro Thr Val Ala Ala Ile Glu Ala Gly Lys Asp Ile
180 185 190

Ala Leu Ala Asn Lys Glu Thr Met Ile Ala Gly Ala Pro Phe Val Leu
195 200 205

Pro Leu Ala His Lys His Asn Ile Lys Ile Leu Pro Ala Asp Ser Glu
210 215 220

His Ser Ala Ile Phe Gln Ser Ile Gln Gly Leu Pro Lys Gly Ala Leu
225 230 235 240

Arg Lys Ile Leu Leu Thr Gly Ser Gly Gly Ala Phe Arg Glu Trp Pro
245 250 255

Ala Glu Lys Met Lys Asp Ile Lys Leu Ala Asp Ala Leu Lys His Pro
260 265 270

Ile Trp Ser Leu Gly Arg Lys Ile Thr Ile Asp Ser Ala Thr Leu Phe
275 280 285

Asn Lys Gly Leu Glu Val Ile Glu Ala His Tyr Leu Phe Gly Ala Ser
290 295 300

Tyr Asp Asp Ile Glu Ile Val Ile His Pro Gln Ser Ile Ile His Ser
305 310 315 320

Leu Val Glu Thr Xaa Asp Ser Ser Val Asn Ala Gln Leu Gly Ile Pro
325 330 335

Asp Met Arg Leu Pro Leu Leu Tyr Thr Leu Ser Trp Pro Glu Arg Ile
340 345 350

Tyr Cys Ser Glu Val Thr Trp Pro Arg Leu Asp Leu Ser Lys Tyr Gly
355 360 365

Ser Leu Thr Phe Tyr Ala Pro Asp Asp Lys Lys Phe Pro Ser Val Asn
370 375 380

Leu Cys Tyr Ala Ala Gly Arg Ala Gly Gly Thr Met Thr Gly Val Leu
385 390 395 400

Ser Ala Ala Asn Glu Lys Ala Val Glu Met Phe Val Glu Glu Lys Ile
405 410 415

Ser Tyr Leu Asp Ile Phe Lys Val Val Glu Leu Thr Cys Gln Glu His
420 425 430

Gln Lys Glu Leu Val Ala Ser Pro Ser Leu Glu Glu Ile Ile His Tyr
435 440 445

Asp Gln Trp Ala Arg Gln Tyr Ala Ala Ser Leu Gln Lys Xaa Phe Lys
450 455 460

Cys Leu Asn Pro Ile Phe Leu Thr Tyr Phe Arg Ser Trp Gly Cys Gly
465 470 475 480

Gly Leu Leu Ala Thr Ala Ser Ile Phe Cys Lys Cys Ile Val Gly Ser
485 490 495

Ser Ile Leu

<210> 9
<211> 784
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (55)

<220>
<221> unsure
<222> (100)

<220>
<221> unsure
<222> (109)

<220>
<221> unsure
<222> (120)

<220>
<221> unsure
<222> (659)

<220>
<221> unsure
<222> (675)

<220>
<221> unsure
<222> (721)

<220>
<221> unsure
<222> (735)

<220>
<221> unsure
<222> (740)

<220>
<221> unsure
<222> (743)

<220>
<221> unsure
<222> (756)

<220>
<221> unsure
<222> (772)

<220>
<221> unsure
<222> (779)..(780)

<400> 9
gcacgggttt attgctcagt gcaggcaaca ccaccaccac cagcctggcc gggangagcg 60
gttccggaac aaggtcgcaa gacttgggat ggaccaaaan ccatttcant tgtggggagn 120
actggttcaa ttggaactca gacactagat attgtggcag agaatccaga taagtttaaa 180
gttgtggcac ttgcagctgg ttcaaatgtt actcttcttg cagaccaggt aaaaagattt 240
aagcctcaac ttgttgctgt tagaaatgag tccctaattg ctgaacttga agaggccttg 300
catgatgttg aagaaaaacc tgagatcatc cctggagagc agggaatcat tgaggttgct 360
cgtcacccag atgcagttag tgtagtcaca ggaatagtag gctgtgcagg actgaagcca 420
acagttgcag cgatagaagc agggaaagac atagctttgg ccaacaaaga gacattgatt 480
gctggaggtc ctttgttctc ctcttgctca gaagcataat gtaaaaatac ttccagctga 540
ttcagaacat ctgccatctt tcagtgtatc caggggttgc cagagggtgc acttaggaga 600
gttattttaa ctgcatctgg aggtgctttc aggggatggc cagttggata actgaagang 660
ttaaagttgc tgatncatta aaacatccta ctggaatatg ggggaaagaa ctgtggactc 720
ngcaaccttt taaanaaggn canaagtaaa tgagcncata ctgtttgggg cngctaagnn 780
catt 784

<210> 10
<211> 215
<212> PRT
<213> Glycine max

<220>
<221> UNSURE
<222> (19)

<220>
<221> UNSURE
<222> (183)

<400> 10
```
Ala Arg Val Tyr Cys Ser Val Gln Ala Thr Pro Pro Pro Pro Ala Trp
 1               5                  10                  15

Pro Gly Xaa Ala Val Pro Glu Gln Gly Arg Lys Thr Trp Asp Gly Pro
            20                  25                  30

Lys Pro Ile Ser Ile Val Gly Ser Thr Gly Ser Ile Gly Thr Gln Thr
        35                  40                  45

Leu Asp Ile Val Ala Glu Asn Pro Asp Lys Phe Lys Val Val Ala Leu
    50                  55                  60

Ala Ala Gly Ser Asn Val Thr Leu Leu Ala Asp Gln Val Lys Arg Phe
65                  70                  75                  80

Lys Pro Gln Leu Val Ala Val Arg Asn Glu Ser Leu Ile Ala Glu Leu
                85                  90                  95

Glu Glu Ala Leu His Asp Val Glu Glu Lys Pro Glu Ile Ile Pro Gly
            100                 105                 110

Glu Gln Gly Ile Ile Glu Val Ala Arg His Pro Asp Ala Val Ser Val
        115                 120                 125

Val Thr Gly Ile Val Gly Cys Ala Gly Leu Lys Pro Thr Val Ala Ala
    130                 135                 140

Ile Glu Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile
145                 150                 155                 160

Ala Gly Gly Pro Leu Ser Pro Leu Ala Gln Lys His Asn Val Lys Ile
                165                 170                 175

Leu Pro Ala Asp Ser Asp Xaa Ser Ala Ile Phe Gln Cys Ile Gln Gly
            180                 185                 190

Leu Pro Glu Gly Ala Leu Arg Arg Val Ile Leu Thr Ala Ser Gly Gly
        195                 200                 205

Ala Phe Arg Gly Trp Pro Val
    210                 215
```

<210> 11
<211> 642
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (506)

<220>
<221> unsure
<222> (516)

<220>
<221> unsure
<222> (534)

<220>
<221> unsure
<222> (554)

<220>
<221> unsure
<222> (576)

<220>
<221> unsure
<222> (584)

<220>
<221> unsure
<222> (597)

<220>
<221> unsure
<222> (601)

<220>
<221> unsure
<222> (606)

<220>
<221> unsure
<222> (625)

<220>
<221> unsure
<222> (628)

<400> 11
ctccttctcc ctcctcgagc tctcctccgg caccaccagg agcaggaggg gagccgcctt 60
ccgccccgc cagcaccagc gcaaagtcga cttcacatat caaaggaggg acaaaagagc 120
tgcctacctg aggacatgct gctccatgca gcagggccca ccgcccgcct ggccaggccg 180
agccgtcgtg gaacctgaga ggaggtcgtg ggagggcccc aagcccatct ccatcgtcgg 240
ctcaaccggt tccataggaa cacagacatt ggacatcgtt gcggagaacc tgacaagttc 300
ccgggttgtc gcccttgctg ctgggtccaa cgtcactcct ctagctgata aggtgaaaac 360
gttcaaacca aactgggtgg tgttaagaaa cgatccatta cttaacgagc taaaggaagc 420
attaactggt tgtgaaagag atccggatta tccctgggga caagtgcata gaggcgcacc 480
caccccggacc attacatcct tacggnatat aggttncaag atcaacctac attncaacat 540
ttaactggaa aatntgcttt gggaacaaaa accttnccag gtgnccttct ctccctncca 600
naacanattg aaatactctg cgatnaanat ctgatatcat ga 642

<210> 12
<211> 94
<212> PRT
<213> Triticum aestivum

<400> 12
Met Gln Gln Gly Pro Pro Pro Ala Trp Pro Gly Arg Ala Val Val Glu
1 5 10 15

Pro Glu Arg Arg Ser Trp Glu Gly Pro Lys Pro Ile Ser Ile Val Gly
20 25 30

Ser Thr Gly Ser Ile Gly Thr Gln Thr Leu Asp Ile Val Ala Glu Asn
35 40 45

Leu Thr Ser Ser Arg Val Val Ala Leu Ala Ala Gly Ser Asn Val Thr
50 55 60

Pro Leu Ala Asp Lys Val Lys Thr Phe Lys Pro Asn Trp Val Val Leu
65 70 75 80

Arg Asn Asp Pro Leu Leu Asn Glu Leu Lys Glu Ala Leu Thr
85 90

<210> 13
<211> 360
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (295)

<220>
<221> unsure
<222> (299)

<220>
<221> unsure
<222> (313)

<220>
<221> unsure
<222> (338)

<220>
<221> unsure
<222> (352)

<400> 13
catctgtcct ggctcagctg ggatggcctg acatgcggct accaatccta tacaccttgt 60
cttggccaga tagagtctac tgctccgagg tcacctggcc tcggctagat ctttgcaagc 120
tgggctcgct gacattcaaa gctcccgaca acgtgaaata cccatcggta gatctccgcc 180
gtacgcggca gggcgagccg gggcaccat gacgggattt ttgagtgctg ctaatgagaa 240
ggcgtggagc ttgttcatcg acgaaaagat taactacctt ggacatcttc aaggngggng 300
agaatacctt ttnacgccaa ccgcaacaac tgggtganag ctcctcccca angggggggg 360

<210> 14
<211> 93
<212> PRT
<213> Triticum aestivum

<220>
<221> UNSURE
<222> (59)

<400> 14

```
Ser Val Leu Ala Gln Leu Gly Trp Pro Asp Met Arg L u Pro Ile Leu
1               5                   10                  15

Tyr Thr Leu Ser Trp Pro Asp Arg Val Tyr Cys Ser Glu Val Thr Trp
            20                  25                  30

Pro Arg Leu Asp Leu Cys Lys Leu Gly Ser Leu Thr Phe Lys Ala Pro
        35                  40                  45

Asp Asn Val Lys Tyr Pro Ser Val Asp Leu Xaa Xaa Tyr Ala Ala Gly
    50                  55                  60

Arg Ala Gly Gly Thr Met Thr Gly Phe Leu Ser Ala Ala Asn Glu Lys
65                  70                  75                  80

Ala Trp Ser Leu Phe Ile Asp Glu Lys Ile Asn Tyr Leu
                85                  90
```

<210> 15
<211> 1847
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (5)..(9)

<400> 15

```
aatgnnnnna tcaggctgtt acataggggg gcttgcattg tacacaccca acctggccta   60
gcctacccta ctacactcgt gccgattcgg cacgagcagc gacggtcgcc accaccgctc  120
ccctccctct ccccctcctc gcccagcggc aattaccaca gcctccccag caagccggga  180
tggctgcact caaggcatcg ttccggggtg agctcagcgc cgcttccttc ctcgactcca  240
gcaggggacc tctcgtccag cacaaagtgg attttacgtt tcaaaggaag ggcaaacgag  300
ctatttcact gagaaggaca tgctgttcta tgcaacaggc tccaccacca gcatggcctg  360
ggcgagctgt tgctgagcct ggccggaggt catgggatgg cccaaagcct atctcgattg  420
ttggttcaac tggttccata ggaacacaga cattggacat tgttgcggag aatcctgata  480
agttcagagt tgttgctctt gctgctggat ccaatgtcac gcttctagct gatcaggtca  540
aaacattcaa acctaagttg gttgctgtaa gaaacgaatc attagttgat gaattgaaag  600
aagccttggc tgattgcgaa gagaagccag aaattattcc tggggagcaa ggtgtcatag  660
aagttgctcg ccatccagat gcagttacag ttgtcacagg gatagtaggt tgtgcagggc  720
tgaagcctac agttgctgca attgaagctg gtaaagacat agcattggca aacaaagaga  780
cacttattgc aggtggtcct tttgtgcttc cccttgcaca caaacacaaa gtgaaaattc  840
ttccagctga ttctgagcac tctgcaatat ttcagtgtat acaaggcttg tccgaaggtg  900
cacttcgtcg cattattcta actgcatcag gtggtgcttt cagggactgg ccagttgaca  960
ggctgaaaga tgtaaaagtt gctgacgctt taaagcatcc aaactggaat atgggaagga 1020
agatcacagt agattctgct actttattca acaagggttt agaagttatt gaagcacatt 1080
atttatttgg tgctgaatat gatgacattg agattgtgat tcacccacag tctatcatac 1140
actctatggt tgaaacccag gattcatctg tcctagctca gttgggatgg ccagatatgc 1200
ggttaccaat cttatacacc ttatcatggc cagatagaat ctattgctct gaggtcacct 1260
ggccccgtct ggatctttgc aagttgggtt cactgacatt cagagctcca gacaacgtaa 1320
aatacccatc aatggaccta gcctatgcag ctggccgcgc tggggggcacc atgacaggag 1380
tcctgagcgc tgctaatgag aaggccgtgg agttgttcat tgacgagaag attagctacc 1440
tggacatatt caaggtggtg gagcttacat gtaacgcgca tcggaacgag ctggtaacaa 1500
gcccgtcact ggaggagatc gtccattacg atctgtgggc gaggagatat gcagccagtc 1560
tacaaccatc ttctggcctg agccctgtcc ctgcataata ggtcgtcacg acaacgttgt 1620
acagcaggag ttctaagata tgatgtgttt gtggctcctg tttccatgtt caattttcag 1680
gcctccacat gaataaaatg catctattcc atgtgatttc ttttatggat gaagtgtgcg 1740
```

```
aagtcgggtg ggaatcagat gcatcccttt cggtggagtt cttacgtagg gttgagcagc 1800
atttttaaa aaggtttttt tacctctgca aaaaaaaaa aaaaaaa                1847

<210> 16
<211> 472
<212> PRT
<213> Zea mays

<400> 16
Met Ala Ala Leu Lys Ala Ser Phe Arg Gly Glu Leu Ser Ala Ala Ser
1               5                   10                  15

Phe Leu Asp Ser Ser Arg Gly Pro Leu Val Gln His Lys Val Asp Phe
            20                  25                  30

Thr Phe Gln Arg Lys Gly Lys Arg Ala Ile Ser Leu Arg Arg Thr Cys
        35                  40                  45

Cys Ser Met Gln Gln Ala Pro Pro Pro Ala Trp Pro Gly Arg Ala Val
    50                  55                  60

Ala Glu Pro Gly Arg Arg Ser Trp Asp Gly Pro Lys Pro Ile Ser Ile
65                  70                  75                  80

Val Gly Ser Thr Gly Ser Ile Gly Thr Gln Thr Leu Asp Ile Val Ala
                85                  90                  95

Glu Asn Pro Asp Lys Phe Arg Val Val Ala Leu Ala Ala Gly Ser Asn
            100                 105                 110

Val Thr Leu Leu Ala Asp Gln Val Lys Thr Phe Lys Pro Lys Leu Val
        115                 120                 125

Ala Val Arg Asn Glu Ser Leu Val Asp Glu Leu Lys Glu Ala Leu Ala
    130                 135                 140

Asp Cys Glu Glu Lys Pro Glu Ile Ile Pro Gly Glu Gln Gly Val Ile
145                 150                 155                 160

Glu Val Ala Arg His Pro Asp Ala Val Thr Val Val Thr Gly Ile Val
                165                 170                 175

Gly Cys Ala Gly Leu Lys Pro Thr Val Ala Ala Ile Glu Ala Gly Lys
            180                 185                 190

Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile Ala Gly Gly Pro Phe
        195                 200                 205

Val Leu Pro Leu Ala His Lys His Lys Val Lys Ile Leu Pro Ala Asp
    210                 215                 220

Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln Gly Leu Ser Glu Gly
225                 230                 235                 240

Ala Leu Arg Arg Ile Ile Leu Thr Ala Ser Gly Gly Ala Phe Arg Asp
                245                 250                 255

Trp Pro Val Asp Arg Leu Lys Asp Val Lys Val Ala Asp Ala Leu Lys
            260                 265                 270
```

His Pro Asn Trp Asn Met Gly Arg Lys Ile Thr Val Asp Ser Ala Thr
275 280 285

Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala His Tyr Leu Phe Gly
290 295 300

Ala Glu Tyr Asp Asp Ile Glu Ile Val Ile His Pro Gln Ser Ile Ile
305 310 315 320

His Ser Met Val Glu Thr Gln Asp Ser Ser Val Leu Ala Gln Leu Gly
325 330 335

Trp Pro Asp Met Arg Leu Pro Ile Leu Tyr Thr Leu Ser Trp Pro Asp
340 345 350

Arg Ile Tyr Cys Ser Glu Val Thr Trp Pro Arg Leu Asp Leu Cys Lys
355 360 365

Leu Gly Ser Leu Thr Phe Arg Ala Pro Asp Asn Val Lys Tyr Pro Ser
370 375 380

Met Asp Leu Ala Tyr Ala Ala Gly Arg Ala Gly Gly Thr Met Thr Gly
385 390 395 400

Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu Leu Phe Ile Asp Glu
405 410 415

Lys Ile Ser Tyr Leu Asp Ile Phe Lys Val Val Glu Leu Thr Cys Asn
420 425 430

Ala His Arg Asn Glu Leu Val Thr Ser Pro Ser Leu Glu Glu Ile Val
435 440 445

His Tyr Asp Leu Trp Ala Arg Arg Tyr Ala Ala Ser Leu Gln Pro Ser
450 455 460

Ser Gly Leu Ser Pro Val Pro Ala
465 470

<210> 17
<211> 2019
<212> DNA
<213> Glycine max

<400> 17
gcagccacca ttattgttgt tattggagat ttcaattctt tgtctttcaa actcctcaag 60
ttgggtttat gtgatgatgg ctctcaacat ctcttctcca gctgaagtca agtccatttt 120
tttcgctgat tccttcaagt ctaactgcct cacagcaaaa ttctcaggtg ggtttgcttt 180
taagagaaaa gagcgtagag cagcatctgg aggacgggtt tattgctcag tgcaggcaac 240
accaccacca ccagcctggc cgggacgagc ggttccggaa caaggtcgca agacttggga 300
tggaccaaaa cccatttcaa ttgtggggag tactggttca attggaactc agacactaga 360
tattgtggca gagaatccag ataagtttaa agttgtggca cttgcagctg gttcaaatgt 420
tactcttctt gcagaccagg taaaaagatt taagcctcaa cttgttgctg ttagaaatga 480
gtccctaatt gctgaacttg aagaggcctt gcatgatgtt gaagaaaaac ctgagatcat 540
ccctggagag cagggaatca ttgaggttgc tcgtcaccca gatgcagtta gtgtagtcac 600
aggaatagta ggctgtgcag gactgaagcc aacagttgca gcgatagaag cagggaaaga 660
catagctttg gccaacaaag agacattgat tgctggaggt ccttttgttc ttcctcttgc 720
tcagaagcat aatgtaaaaa tacttccagc tgattcagaa cattctgcca tctttcagtg 780
tatccagggg ttgccagagg gtgcacttag gagagttatt ttaactgcat ctggaggtgc 840
tttcagggat tggccagttg ataaactgaa agatgttaaa gttgctgatg cattaaaaca 900

```
tcctaactgg aatatgggga aaaagataac tgtggactct gctacccttt ttaataaggg  960
tctagaagta attgaagcac attacttgtt tggagctgac tacgatcata ttgagattgt 1020
cattcatcca caatcaatca tacattcaat gattgaaaca caggattcat ctgttcttgc 1080
acaattgggg tggcctgata tgcgtttgcc aatcctctat acattatcat ggcctgacag 1140
gatttattgt tctgaagtca cttggccacg ccttgatctt tgcaagcttg gttcacttac 1200
atttaaaact ccagataatg taaagtatcc atccatgaat cttgcatttt ctgctggccg 1260
tgctggaggc acaatgacag gagttcttag tgcagcaaat gaaaaagctg tagagatgtt 1320
tattgatgaa aagataagct attggaattt attcaaagtt gtggagctaa catgtgagaa 1380
gcatcaaaat gaattggtat cctctccttc ccttgaggaa attattcact atgacctgtg 1440
ggcgcgaaaa tatgctgcta gtctgcaaga ctcttccagc ttcactccta ttcttgcatg 1500
aggatgatta aactagggat gtggctgatg cttcccaatt gcctgctttc accataattt 1560
cttcgggcat tgaacaatgt agaatggtgc attccacaga tgttgaaaat taaataggtt 1620
ttttgtttat ggaatgttgg tgttttaaca cctttcaatt gatcttatag ttttgtcgta 1680
atttcatgga aaacgatgtc tttttaatag tcaataggag cctaggaggt tggttggttg 1740
cctatgaatg tgtcaaagtc aagaagggga atggattttc tcatattcaa aatttacatg 1800
atgtggtcaa ctagaagttt tgtatttctc tttttctaat agaattaaat aggtggagtc 1860
ttacaaaaat taacagagat agacacaaaa gttgaccaat caccaatcac tttcataaaa 1920
ggattccttt tctttttcct cagcacacat tcgttggctg atattattat atgaaattgg 1980
tattatttgg atatcatagc taaaaaaaaa aaaaaaaaa                        2019
```

<210> 18
<211> 475
<212> PRT
<213> Glycine max

<400> 18

```
Met Met Ala Leu Asn Ile Ser Ser Pro Ala Glu Val Lys Ser Ile Phe
1               5                   10                  15

Phe Ala Asp Ser Phe Lys Ser Asn Cys Leu Thr Ala Lys Phe Ser Gly
            20                  25                  30

Gly Phe Ala Phe Lys Arg Lys Glu Arg Arg Ala Ala Ser Gly Gly Arg
        35                  40                  45

Val Tyr Cys Ser Val Gln Ala Thr Pro Pro Pro Pro Ala Trp Pro Gly
    50                  55                  60

Arg Ala Val Pro Glu Gln Gly Arg Lys Thr Trp Asp Gly Pro Lys Pro
65                  70                  75                  80

Ile Ser Ile Val Gly Ser Thr Gly Ser Ile Gly Thr Gln Thr Leu Asp
                85                  90                  95

Ile Val Ala Glu Asn Pro Asp Lys Phe Lys Val Val Ala Leu Ala Ala
            100                 105                 110

Gly Ser Asn Val Thr Leu Leu Ala Asp Gln Val Lys Arg Phe Lys Pro
        115                 120                 125

Gln Leu Val Ala Val Arg Asn Glu Ser Leu Ile Ala Glu Leu Glu Glu
    130                 135                 140

Ala Leu His Asp Val Glu Glu Lys Pro Glu Ile Ile Pro Gly Glu Gln
145                 150                 155                 160

Gly Ile Ile Glu Val Ala Arg His Pro Asp Ala Val Ser Val Val Thr
                165                 170                 175
```

Gly Ile Val Gly Cys Ala Gly Leu Lys Pro Thr Val Ala Ala Ile Glu
180 185 190

Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile Ala Gly
195 200 205

Gly Pro Phe Val Leu Pro Leu Ala Gln Lys His Asn Val Lys Ile Leu
210 215 220

Pro Ala Asp Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln Gly Leu
225 230 235 240

Pro Glu Gly Ala Leu Arg Arg Val Ile Leu Thr Ala Ser Gly Gly Ala
245 250 255

Phe Arg Asp Trp Pro Val Asp Lys Leu Lys Asp Val Lys Val Ala Asp
260 265 270

Ala Leu Lys His Pro Asn Trp Asn Met Gly Lys Lys Ile Thr Val Asp
275 280 285

Ser Ala Thr Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala His Tyr
290 295 300

Leu Phe Gly Ala Asp Tyr Asp His Ile Glu Ile Val Ile His Pro Gln
305 310 315 320

Ser Ile Ile His Ser Met Ile Glu Thr Gln Asp Ser Ser Val Leu Ala
325 330 335

Gln Leu Gly Trp Pro Asp Met Arg Leu Pro Ile Leu Tyr Thr Leu Ser
340 345 350

Trp Pro Asp Arg Ile Tyr Cys Ser Glu Val Thr Trp Pro Arg Leu Asp
355 360 365

Leu Cys Lys Leu Gly Ser Leu Thr Phe Lys Thr Pro Asp Asn Val Lys
370 375 380

Tyr Pro Ser Met Asn Leu Ala Phe Ser Ala Gly Arg Ala Gly Gly Thr
385 390 395 400

Met Thr Gly Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu Met Phe
405 410 415

Ile Asp Glu Lys Ile Ser Tyr Trp Asn Leu Phe Lys Val Val Glu Leu
420 425 430

Thr Cys Glu Lys His Gln Asn Glu Leu Val Ser Ser Pro Ser Leu Glu
435 440 445

Glu Ile Ile His Tyr Asp Leu Trp Ala Arg Lys Tyr Ala Ala Ser Leu
450 455 460

Gln Asp Ser Ser Ser Phe Thr Pro Ile Leu Ala
465 470 475

<210> 19
<211> 1640